How to construct Detailed Tables from the ACS Summary Files

This document explains the process for constructing US Census American Community Survey (ACS) Detailed Tables, as implemented in the Python program censusACS.py.

Refer to the Census Bureau's ACS Summary File Technical Documentation (ACS_Summary_File_Technical_Documentation.pdf) for further details.

Downloads

Download these files from the US Census website, either from your browser starting at https://www.census.gov/programs-surveys/acs/, or programmatically from https://www2.census.gov/programs-surveys/acs/summary_file (example based on 2015 ACS 5-year Estimates data):

- ACS_2015_SF_5YR_Appendices.xls

- Colorado_Tracts_Block_Groups_Only.zip (or other state)

- 2015_5yr_Summary_FileTemplates.zip

ACS_2015_SF_5YR_Appendices.xls

Appendix A in file ACS_2015_SF_5YR_Appendices.xls contains ACS Table names, titles, geographic restrictions, Summary File sequence numbers, and Summary File start/end position numbers. I explain the sequence and start/end numbers below.

Table Number	Table Title	Geography Restrictions	Summary File Sequence Number	Summary File Starting and Ending Positions
B00001	Unweighted Sample Count Of The Population		0001	7-7
B00002	Unweighted Sample Housing Units		0001	8-8
B01001	Sex By Age		0002	7-55
B01001A	Sex By Age (White Alone)	No Blockgroups	0002	56-86
B01001B	Sex By Age (Black Or African American Alone)	No Blockgroups	0002	87-117

In brief, Appendix A refers to the necessary estimate and margin-of-error summary files, and corresponding data columns within each file, for each table.

Appendix B is unnecessary for our purposes.
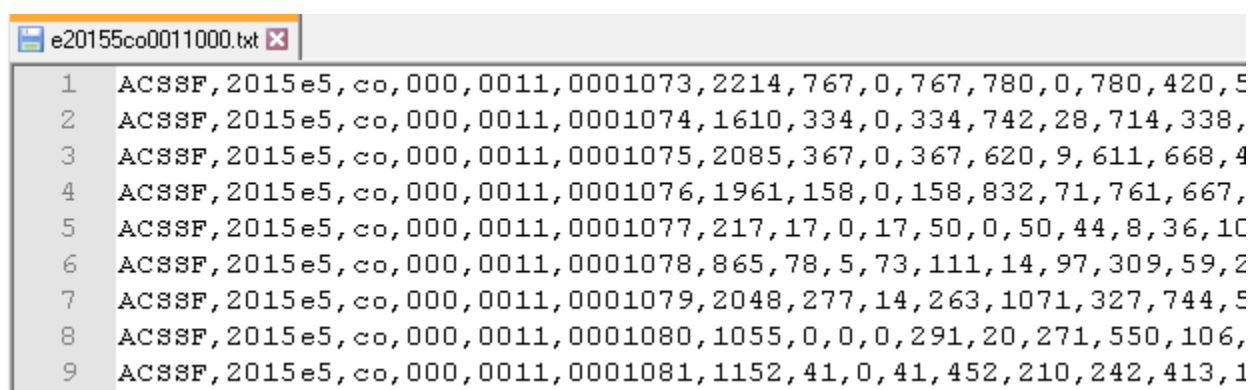
Colorado_Tracts_Block_Groups_Only.zip

The state-level <State>_ Tracts_Block_Groups_Only.zip file is a multi-file archive containing a geographic reference file (available as .txt and .csv), the sequenced estimate files, and the sequenced margin-of-error files. By sequenced, I mean a series of files matching the sequence file numbers referred to by the "Summary File Sequence Number" column in Appendix A.

Naming convention for sequenced estimate and margin-of-error files:

Example: Estimate file e20155co0001000.txt (e 2015 5 co 0001 000.txt)

Example	Name	Range or Type
e	File Type	e=estimate, m=margin of error
2015	Reference Year	ACS data year (last year of multiyear period)
5	Period Covered	1=1-year, 5=5-year
co	State Level	US or abbreviations for state, DC and Puerto Rico
0001	Sequence Number	0001 to 9999
000	Reserved	Iteration value reserved for future use

The estimate and margin-of-error files contain unlabeled survey data necessary to construct each detailed table.

Information from other sources — summary file sequence numbers, start/end positions, and logical record numbers — provide the necessary pointers to extract the correct data from the estimate and margin-of-error files.

The geography file (choose either the fixed-width g20155co.txt or comma-separated g20155co.csv file) provides the Logical Record Numbers needed to locate the desired rows within the estimate and margins files, as well as Geographic Identifiers to associate each record with its geographic entity (e.g. Block Group).

2015_5yr_Summary_FileTemplates.zip

The "template" files contained in 2015_5yr_Summary_FileTemplates.zip provide descriptive column names to accompany each summary file (geography, estimate, or margins) described above. File 2015_5yr_Summary_FileTemplates.zip is a multi-file archive containing one geography file template (2015_SFGeoFileTemplate.xls) and 122 sequence file templates (Seq1.xls - Seq122.xls), the latter matching with the sequenced estimate and margins data files.

In brief, the summary file templates provide descriptive column names for the generated output tables.

Step-by-step example for creating Table B01002, "Median Age By Sex," for all Colorado Census Block Groups.

1.   From Appendix A in file ACS_2015_SF_5YR_Appendices.xls, find the row containing table B01002 in the "Table Number" column.

     Note: Occasionally, but rarely, a table spans multiple sequence files, so be sure to look for all rows for a given table number.

2.   Record the row information found in Step #1 — Table Number (B01002), Table Title (Median Age By Sex), Geography Restrictions (none), Summary File Sequence Number (0003), Summary File Starting and Ending Positions (100-102) — for later use.

3.   From geography file g20155co.csv in Summary File Colorado_Tracts_Block_Groups_Only.zip, filter for and save the 3,532 rows containing Block Group summary level value (150) in the third column.

4.   From the geography file template 2015_SFGeoFileTemplate.xls in 2015_5yr_Summary_FileTemplates.zip, use Row #2 for the descriptive column names to associate with the geo data collected in Step #3 above.

5.   Using the "Logical Record Number" and "Geographic Identifier" columns for the block group data from Step #3, form a new table with these columns only.

6.   Referencing the Summary File Sequence Number 0003 from Step #2, open files e20155co0003000.txt and m20155co0003000.txt from the Summary File archive Colorado_Tracts_Block_Groups_Only.zip.

7.   Use the logical record numbers from the table constructed in Step #5 to locate the corresponding rows (matching column 6) in the estimate and margins files opened in Step #6.

8.   For each row in the estimate and margins files, obtain survey data values from the columns between the Summary File Starting and Ending Positions (inclusive) obtained from Appendix A in Step #2.

9.   Use the appropriate template file (e.g. Seq3.xls) from file 2015_5yr_Summary_FileTemplates.zip to get descriptive column names to associate with the survey data collected in Step #8.

10. Merge by Logical Record Number the tables created in Step #5 and Step #9, then form a final output table by retaining only the Geographic Identifier column, estimate columns, and margin-of-error columns.

11. Save the output table to a CSV file or database.